A Major Project Abstract

On

AUTOMATED IMAGE DESCRIPTION GENERATOR USING DEEP LEARNING

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted By

K. PAVANI SUREKHA	**(208R1A05E5)**
CHILUKURI NITHISH KUMAR	**(208R1A05D3)**
MANEPALLI VENKAT SAI	**(208R1A05F7)**
VENNAPU RITHESH CHANDRA	**(208R1A05H7)**

Under the guidance of

Mrs.Sumera Jabeen

Assistant Professor, Department of CSE

Department of Computer Science & Engineering

CMR ENGINEERING COLLEGE

UGC AUTONOMOUS

(Accredited by NBA Approved by AICTE,NEW DELHI, Affiliated to JNTU Hyderabad) Kandlakoya, Medchal Road, Medchal Malkajgiri Dist. Hyderabad-501 401)

(2023-2024)

ABSTRACT

Nowadays, an image caption generator has become the need of the hour, be it for social media enthusiasts or visually impaired people. It can be used as a plugin in currently trending social media platforms to recommend suitable captions for people to attach to their post or can be used by visually impaired people to understand the image content on the web thus eradicating any ambiguity in image meaning in turn also free of any discrepancy in knowledge acquisition. The proposed paper aims to generate a description of an image also called as image captioning, using CNN-LSTM architecture such that CNN layers will help in extraction of the input data and LSTM will extract relevant information throughout the processing of input such that the current word acts as an input for the prediction of the next word. The programming language used will be Python 3 and machine learning techniques. This paper will also elaborate on the functions and structure of the various Neural networks involved.

EXISTING SYSTEM

The goal of this model is to generate descriptions of image regions. During training, the input to the model is a set of images and their corresponding sentence descriptions. They first present a model that aligns sentence snippets to the visual regions that they describe through a multimodal embedding. And then treat these correspondences as training data for a second, multimodal Recurrent Neural Network model that learns to generate the snippets. Learning to align visual and language data this alignment model assumes an input dataset of images and their sentence descriptions. Their key insight is that sentences written by people make frequent references to some particular, but unknown location in the image. It would like to infer these latent correspondences, with the eventual goal of later learning to generate these snippets from image regions. They build on the approach, which learn to ground dependency tree relations to image regions with a ranking objective. Their contribution is in the use of bidirectional recurrent neural network to compute word representations in the sentence, dispensing of the need to compute dependency trees and allowing unbounded interactions of words and their context in the sentence. They also substantially simplify their objective and show that both modifications improve ranking performance. They have described the transformations that map every image and sentence into a set of vectors in a common h-dimensional space. Since the supervision is at the level of entire images and sentences, our strategy is to formulate an image-sentence score as a function of the individual region word scores. Intuitively, a sentence-image pair should have a high matching score if its words have a confident support in the image.

DISADVANTAGES

- An existing system training time is very high.
- The image caption predicting accuracy of existing system is very poor.
- An existing system preprocessor takes more time for image transformation.

PROPOSED SYSTEM

Our model, uses a convolutional neural network to generate a dense feature vector from an input image. This dense vector, also known as an embedding, creates appropriate captions for the images that are supplied as an output and can be utilised as an input into other algorithms. We are appending start and end words to each and every sentence at the beginning and the end. While training the model we have generated five captions to each and every image. These embedding forms a way to represent the underlying image, which is further utilized to generate appropriate captions associated with the image.

Here, the proposed model is based on the neural network which utilizes probability concept to determine chance of occurrence of a favorable event. The underlying mathematical model is optimized and trained to obtain most suitable outcome. This is achieved after multiple iterations and the probability of an appropriate caption is maximized.

CNN

Conventional neural networks (CNNs) are the neural networks which are mostly in the form of a matrix for the input images. CNN makes use of dividing the cluster in to multiple frames and recognizes using the training. The model can differentiate between objects based on size For example, a bird or a plane. The proposed methodology uses the commonly used scanning process to scan the object under test in horizontal, left to right, and vertical, from top to bottom. It can handle images that have been translated, flipped, scaled, and have had their colors changed.

LSTM(Long short-term memory)

These belong to RNN networks, which are proficient at foreseeing sequences. Based on the paragraph that came before it, we can guess what the following phrases will be By addressing RNN's shortcomings, it has been discovered to be more effective than traditional RNN. LSTM can filter out unnecessary input and keep track of pertinent data throughout processing. In terms of overcoming the shortcomings of RNNS with short term memory, it has done better than ordinary RNNs.

ADVANTAGES

- By this system, we get high accurated output prediction for image captioning.
- The training time of this system compared to existing system is very less.
- The system takes very less time to predict and generate output.

SYSTEM REQUIREMENTS :-

❖ H/W SYSTEM CONFIGURATION:-

- Processor - intel processor 64- bit
- RAM - 8 GB (min)
- Hard Disk - 20 GB
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - SVGA

❖ SOFTWARE REQUIREMENTS:-

- Operating system : Windows 7 and above
- Coding Language : Python.
- Front-End : Python.
- Back-End : Django-ORM
- Designing : HTML, CSS, JAVASCRIPT.
- Editor : Jupyter
- Data Base : MySQL (WAMP Server).

Mini Project Coordinator **Internal Guide**